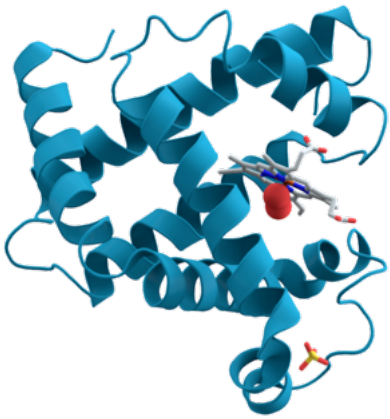
Protein

Proteins are large biomolecules and macromolecules that comprise one or more long chains of amino acid residues. Proteins perform a vast array of functions within organisms, including catalysing metabolic reactions, DNA replication, responding to stimuli, providing structure to cells and organisms, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in protein folding into a specific 3D structure that determines its activity.



A representation of the 3D structure of the protein myoglobin showing turquoise α-helices. This protein was the first to have its structure solved by X-ray crystallography. Toward the right-center among the coils, a prosthetic group called a heme group (shown in gray) with a bound oxygen molecule (red).

A linear chain of amino acid residues is called a polypeptide. A protein contains at least one long polypeptide. Short polypeptides, containing less than 20–30 residues, are rarely considered to be proteins and are commonly called peptides. The individual amino acid residues are bonded together by peptide bonds and adjacent amino acid residues. The sequence of amino acid residues in a protein is defined by the sequence of a gene, which is encoded in the genetic code. In general, the genetic code specifies 20 standard amino acids; but in certain organisms the genetic code can include selenocysteine and—in certain archaea—pyrrolysine. Shortly after or even during synthesis, the residues in a protein are often chemically modified by post-translational modification, which alters the physical and chemical properties, folding, stability,

activity, and ultimately, the function of the proteins. Some proteins have non-peptide groups attached, which can be called prosthetic groups or cofactors. Proteins can also work together to achieve a particular function, and they often associate to form stable protein complexes.

Once formed, proteins only exist for a certain period and are then degraded and recycled by the cell's machinery through the process of protein turnover. A protein's lifespan is measured in terms of its half-life and covers a wide range. They can exist for minutes or years with an average lifespan of 1–2 days in mammalian cells. Abnormal or misfolded proteins are degraded more rapidly either due to being targeted for destruction or due to being unstable.

Like other biological macromolecules such as polysaccharides and nucleic acids, proteins are essential parts of organisms and participate in virtually every process within cells. Many proteins are enzymes that catalyse biochemical reactions and are vital to metabolism. Proteins also have structural or mechanical functions, such as actin and myosin in muscle and the proteins in the cytoskeleton, which form a system of scaffolding that maintains cell shape. Other proteins are important in cell signaling, immune responses, cell adhesion, and the cell cycle. In animals, proteins are needed in the diet to provide the essential amino acids that cannot be synthesized. Digestion breaks the proteins down for metabolic use.

Proteins may be purified from other cellular components using a variety of techniques such as ultracentrifugation, precipitation, electrophoresis, and chromatography; the advent of genetic engineering has made possible a number of methods to facilitate purification. Methods commonly used to study protein structure and function include immunohistochemistry, site-directed mutagenesis, X-ray crystallography, nuclear magnetic resonance and mass spectrometry.

History and etymology

Proteins were recognized as a distinct class of biological molecules in the eighteenth century by Antoine Fourcroy and others, distinguished by the molecules' ability to coagulate or flocculate under treatments with heat or acid.[1] Noted examples at the time included albumin from egg whites, blood serum albumin, fibrin, and wheat gluten.

Proteins were first described by the Dutch chemist Gerardus Johannes Mulder and named by the Swedish chemist Jöns Jacob Berzelius in 1838.[2][3] Mulder carried out elemental analysis of common proteins and found that nearly all proteins had the same empirical formula, $C_{400}H_{620}N_{100}O_{120}P_1S_1$.[4] He came to the erroneous conclusion that they might be composed of

a single type of (very large) molecule. The term "protein" to describe these molecules was proposed by Mulder's associate Berzelius; protein is derived from the Greek word πρώτειος (*proteios*), meaning "primary",[5] "in the lead", or "standing in front",[6] + *-in*. Mulder went on to identify the products of protein degradation such as the amino acid leucine for which he found a (nearly correct) molecular weight of 131 Da.[4] Prior to "protein", other names were used, like "albumins" or "albuminous materials" (*Eiweisskörper*, in German).[7]

Early nutritional scientists such as the German Carl von Voit believed that protein was the most important nutrient for maintaining the structure of the body, because it was generally believed that "flesh makes flesh."[8] Karl Heinrich Ritthausen extended known protein forms with the identification of glutamic acid. At the Connecticut Agricultural Experiment Station a detailed review of the vegetable proteins was compiled by Thomas Burr Osborne. Working with Lafayette Mendel and applying Liebig's law of the minimum in feeding laboratory rats, the nutritionally essential amino acids were established. The work was continued and communicated by William Cumming Rose. The understanding of proteins as polypeptides came through the work of Franz Hofmeister and Hermann Emil Fischer in 1902.[9][10] The central role of proteins as enzymes in living organisms was not fully appreciated until 1926, when James B. Sumner showed that the enzyme urease was in fact a protein.[11]

The difficulty in purifying proteins in large quantities made them very difficult for early protein biochemists to study. Hence, early studies focused on proteins that could be purified in large quantities, e.g., those of blood, egg white, various toxins, and digestive/metabolic enzymes obtained from slaughterhouses. In the 1950s, the Armour Hot Dog Co. purified 1 kg of pure bovine pancreatic ribonuclease A and made it freely available to scientists; this gesture helped ribonuclease A become a major target for biochemical study for the following decades.[4]

Linus Pauling is credited with the successful prediction of regular protein secondary structures based on hydrogen bonding, an idea first put forth by William Astbury in 1933.[12] Later work by Walter Kauzmann on denaturation,[13][14] based partly on previous studies by Kaj Linderstrøm-Lang,[15] contributed an understanding of protein folding and structure mediated by hydrophobic interactions.

The first protein to be sequenced was insulin, by Frederick Sanger, in 1949. Sanger correctly determined the amino acid sequence of insulin, thus conclusively demonstrating that proteins consisted of linear polymers of amino acids rather than branched chains, colloids, or cyclols.[16] He won the Nobel Prize for this achievement in 1958.[17]

John Kendrew with model of myoglobin in progress

With the development of X-ray crystallography, it became possible to sequence protein structures.[18] The first protein structures to be solved were hemoglobin by Max Perutz and myoglobin by John Kendrew, in 1958.[19][20] The use of computers and increasing computing power also supported the sequencing of complex proteins. In 1999, Roger Kornberg succeeded in sequencing the highly complex structure of RNA polymerase using high intensity X-rays from synchrotrons.[18]

Since then, cryo-electron microscopy (cryo-EM) of large macromolecular assemblies[21] has been developed. Cryo-EM uses protein samples that are frozen rather than crystals, and beams of electrons rather than x-rays. It causes less damage to the sample, allowing scientists to obtain more information and analyze larger structures.[18] Computational protein structure prediction of small protein domains[22] has also helped researchers to approach atomic-level resolution of protein structures. As of 2017, the Protein Data Bank has over 126,060 atomic-resolution structures of proteins.[23]

Number of proteins encoded in genomes

The number of proteins encoded in a genome roughly corresponds to the number of genes (although there may be a significant number of genes that encode RNA of protein, e.g. ribosomal RNAs). Viruses typically encode a few to a few hundred proteins, archaea and bacteria a few hundred to a few thousand, while eukaryotes typically encode a few thousand up to tens of thousands of proteins (see genome size for a list of examples).

Classification

Proteins are primarily classified by sequence and structure, although other classifications are commonly used. Especially for enzymes the EC number system provides a functional classification scheme. Similarly, the gene ontology classifies both genes and proteins by their

biological and biochemical function, but also by their intracellular location.

Sequence similarity is used to classify proteins both in terms of evolutionary and functional similarity. This may use either whole proteins or protein domains, especially in multi-domain proteins. Protein domains allow protein classification by a combination of sequence, structure and function, and thy can be combined in many different ways. In an early study of 170,000 proteins, about two-thirds were assigned at least one domain, with larger proteins containing more domains (e.g. proteins larger than 600 amino acids having an average of more than 5 domains).[24]

Biochemistry

Chemical structure of the peptide bond (bottom) and the three-dimensional structure of a peptide bond between an alanine and an adjacent amino acid (top/inset). The bond itself is made of the CHON elements.

Resonance structures of the peptide bond that links individual amino acids to form a protein polymer

Most proteins consist of linear polymers built from series of up to 20 different L-α- amino acids. All proteinogenic amino acids possess common structural features, including an α-carbon to which an amino group, a carboxyl group, and a variable side chain are bonded. Only proline differs from this basic structure as it contains an unusual ring to the N-end amine group, which forces the CO–NH amide moiety into a fixed conformation.[25] The side chains of the standard amino acids, detailed in the list of standard amino acids, have a great variety of chemical structures and properties; it is the combined effect of all of the amino acid side chains in a

protein that ultimately determines its three-dimensional structure and its chemical reactivity.[26] The amino acids in a polypeptide chain are linked by peptide bonds. Once linked in the protein chain, an individual amino acid is called a *residue,* and the linked series of carbon, nitrogen, and oxygen atoms are known as the *main chain* or *protein backbone*.[27]:19

The peptide bond has two resonance forms that contribute some double-bond character and inhibit rotation around its axis, so that the alpha carbons are roughly coplanar. The other two dihedral angles in the peptide bond determine the local shape assumed by the protein backbone.[27]:31 The end with a free amino group is known as the N-terminus or amino terminus, whereas the end of the protein with a free carboxyl group is known as the C-terminus or carboxy terminus (the sequence of the protein is written from N-terminus to C-terminus, from left to right).

The words *protein*, *polypeptide,* and *peptide* are a little ambiguous and can overlap in meaning. *Protein* is generally used to refer to the complete biological molecule in a stable conformation, whereas *peptide* is generally reserved for a short amino acid oligomers often lacking a stable 3D structure. But the boundary between the two is not well defined and usually lies near 20–30 residues.[28] *Polypeptide* can refer to any single linear chain of amino acids, usually regardless of length, but often implies an absence of a defined conformation.

Interactions

Proteins can interact with many types of molecules, including with other proteins, with lipids, with carbohydrates, and with DNA.[29][30][27][31]

Abundance in cells

It has been estimated that average-sized bacteria contain about 2 million proteins per cell (e.g. *E. coli* and *Staphylococcus aureus*). Smaller bacteria, such as *Mycoplasma* or *spirochetes* contain fewer molecules, on the order of 50,000 to 1 million. By contrast, eukaryotic cells are larger and thus contain much more protein. For instance, yeast cells have been estimated to contain about 50 million proteins and human cells on the order of 1 to 3 billion.[32] The concentration of individual protein copies ranges from a few molecules per cell up to 20 million.[33] Not all genes coding proteins are expressed in most cells and their number depends on, for example, cell type and external stimuli. For instance, of the 20,000 or so proteins encoded by the human genome, only 6,000 are detected in lymphoblastoid cells.[34]

Synthesis

Biosynthesis

A ribosome produces a protein using mRNA as template

The DNA sequence of a gene encodes the amino acid sequence of a protein

Proteins are assembled from amino acids using information encoded in genes. Each protein has its own unique amino acid sequence that is specified by the nucleotide sequence of the gene encoding this protein. The genetic code is a set of three-nucleotide sets called codons and each three-nucleotide combination designates an amino acid, for example AUG (adenine–uracil–guanine) is the code for methionine. Because DNA contains four nucleotides, the total number of possible codons is 64; hence, there is some redundancy in the genetic code, with some amino acids specified by more than one codon.[31]:1002–42 Genes encoded in DNA are first transcribed into pre-messenger RNA (mRNA) by proteins such as RNA polymerase. Most organisms then process the pre-mRNA (also known as a *primary transcript*) using various forms of post-transcriptional modification to form the mature mRNA, which is then used as a template for protein synthesis by the ribosome. In prokaryotes the mRNA may either be used as soon as it is produced, or be bound by a ribosome after having moved away from the nucleoid. In contrast, eukaryotes make mRNA in the cell nucleus and then translocate it across the nuclear membrane into the cytoplasm, where protein synthesis then takes place. The rate of protein synthesis is higher in prokaryotes than eukaryotes and can reach up to 20 amino acids per second.[35]

The process of synthesizing a protein from an mRNA template is known as translation. The mRNA is loaded onto the ribosome and is read three nucleotides at a time by matching each

codon to its base pairing anticodon located on a transfer RNA molecule, which carries the amino acid corresponding to the codon it recognizes. The enzyme aminoacyl tRNA synthetase "charges" the tRNA molecules with the correct amino acids. The growing polypeptide is often termed the *nascent chain*. Proteins are always biosynthesized from N-terminus to C-terminus.[31]:1002–42

The size of a synthesized protein can be measured by the number of amino acids it contains and by its total molecular mass, which is normally reported in units of *daltons* (synonymous with atomic mass units), or the derivative unit kilodalton (kDa). The average size of a protein increases from Archaea to Bacteria to Eukaryote (283, 311, 438 residues and 31, 34, 49 kDa respectively) due to a bigger number of protein domains constituting proteins in higher organisms.[36] For instance, yeast proteins are on average 466 amino acids long and 53 kDa in mass.[28] The largest known proteins are the titins, a component of the muscle sarcomere, with a molecular mass of almost 3,000 kDa and a total length of almost 27,000 amino acids.[37]

Chemical synthesis

Short proteins can also be synthesized chemically by a family of methods known as peptide synthesis, which rely on organic synthesis techniques such as chemical ligation to produce peptides in high yield.[38] Chemical synthesis allows for the introduction of non-natural amino acids into polypeptide chains, such as attachment of fluorescent probes to amino acid side chains.[39] These methods are useful in laboratory biochemistry and cell biology, though generally not for commercial applications. Chemical synthesis is inefficient for polypeptides longer than about 300 amino acids, and the synthesized proteins may not readily assume their native tertiary structure. Most chemical synthesis methods proceed from C-terminus to N-terminus, opposite the biological reaction.[40]

Structure

The crystal structure of the chaperonin, a huge protein complex. A single protein subunit is highlighted. Chaperonins assist protein folding.

Three possible representations of the three-dimensional structure of the protein triose phosphate isomerase. **Left**: All-atom representation colored by atom type. **Middle:** Simplified representation illustrating the backbone conformation, colored by secondary structure. **Right**: Solvent-accessible surface representation colored by residue type (acidic residues red, basic residues blue, polar residues green, nonpolar residues white).

Most proteins fold into unique 3D structures. The shape into which a protein naturally folds is known as its native conformation.[27]:36 Although many proteins can fold unassisted, simply through the chemical properties of their amino acids, others require the aid of molecular chaperones to fold into their native states.[27]:37 Biochemists often refer to four distinct aspects of a protein's structure:[27]:30–34

- *Primary structure*: the amino acid sequence. A protein is a polyamide.
- *Secondary structure*: regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the α-helix, β-sheet and turns. Because secondary structures are local, many regions of different secondary structure can be present in the same protein molecule.
- *Tertiary structure*: the overall shape of a single protein molecule; the spatial relationship of the secondary structures to one another. Tertiary structure is generally stabilized by nonlocal interactions, most commonly the formation of a hydrophobic core, but also through salt bridges, hydrogen bonds, disulfide bonds, and even post-translational modifications. The term "tertiary structure" is often used as synonymous with the term *fold*. The tertiary structure is what controls the basic function of the protein.
- *Quaternary structure*: the structure formed by several protein molecules (polypeptide chains), usually called *protein subunits* in this context, which function as a single protein complex.
- *Quinary structure*: the signatures of protein surface that organize the crowded cellular interior. Quinary structure is dependent on transient, yet essential, macromolecular interactions that occur inside living cells.

Proteins are not entirely rigid molecules. In addition to these levels of structure, proteins may shift between several related structures while they perform their functions. In the context of these functional rearrangements, these tertiary or quaternary structures are usually referred to as "conformations", and transitions between them are called *conformational changes*. Such changes are often induced by the binding of a substrate molecule to an enzyme's active site, or the physical region of the protein that participates in chemical catalysis. In solution, proteins also undergo variation in structure through thermal vibration and the collision with other molecules.[31]:368–75

Molecular surface of several proteins showing their comparative sizes. From left to right are: immunoglobulin G (IgG, an antibody), hemoglobin, insulin (a hormone), adenylate kinase (an enzyme), and glutamine synthetase (an enzyme).

Proteins can be informally divided into three main classes, which correlate with typical tertiary structures: globular proteins, fibrous proteins, and membrane proteins. Almost all globular proteins are soluble and many are enzymes. Fibrous proteins are often structural, such as collagen, the major component of connective tissue, or keratin, the protein component of hair and nails. Membrane proteins often serve as receptors or provide channels for polar or charged molecules to pass through the cell membrane.[31]:165–85

A special case of intramolecular hydrogen bonds within proteins, poorly shielded from water attack and hence promoting their own dehydration, are called dehydrons.[41]

Protein domains

Many proteins are composed of several protein domains, i.e. segments of a protein that fold into distinct structural units. Domains usually also have specific functions, such as enzymatic activities (e.g. kinase) or they serve as binding modules (e.g. the SH3 domain binds to proline-rich sequences in other proteins).

Sequence motif

Short amino acid sequences within proteins often act as recognition sites for other proteins.[42] For instance, SH3 domains typically bind to short PxxP motifs (i.e. 2 prolines [P], separated by two unspecified amino acids [x], although the surrounding amino acids may determine the exact binding specificity). Many such motifs has been collected in the Eukaryotic Linear Motif (ELM) database.

Protein topology

Topology of a protein describes the entanglement of the backbone and the arrangement of contacts within the folded chain.[43] Two theoretical frameworks of knot theory and Circuit topology have been applied to characterise protein topology. Being able to describe protein topology opens up new pathways for protein engineering and pharmaceutical development, and adds to our understanding of protein misfolding diseases such as neuromuscular disorders and cancer.

Cellular functions

Proteins are the chief actors within the cell, said to be carrying out the duties specified by the information encoded in genes.[28] With the exception of certain types of RNA, most other biological molecules are relatively inert elements upon which proteins act. Proteins make up half the dry weight of an *Escherichia coli* cell, whereas other macromolecules such as DNA and RNA make up only 3% and 20%, respectively.[44] The set of proteins expressed in a particular cell or cell type is known as its proteome.

The enzyme hexokinase is shown as a conventional ball-and-stick molecular model. To scale in the top right-hand corner are two of its substrates, ATP and glucose.

The chief characteristic of proteins that also allows their diverse set of functions is their ability to bind other molecules specifically and tightly. The region of the protein responsible for binding

another molecule is known as the binding site and is often a depression or "pocket" on the molecular surface. This binding ability is mediated by the tertiary structure of the protein, which defines the binding site pocket, and by the chemical properties of the surrounding amino acids' side chains. Protein binding can be extraordinarily tight and specific; for example, the ribonuclease inhibitor protein binds to human angiogenin with a sub-femtomolar dissociation constant (<10^{-15} M) but does not bind at all to its amphibian homolog onconase (> 1 M). Extremely minor chemical changes such as the addition of a single methyl group to a binding partner can sometimes suffice to nearly eliminate binding; for example, the aminoacyl tRNA synthetase specific to the amino acid valine discriminates against the very similar side chain of the amino acid isoleucine.[45]

Proteins can bind to other proteins as well as to small-molecule substrates. When proteins bind specifically to other copies of the same molecule, they can oligomerize to form fibrils; this process occurs often in structural proteins that consist of globular monomers that self-associate to form rigid fibers. Protein–protein interactions also regulate enzymatic activity, control progression through the cell cycle, and allow the assembly of large protein complexes that carry out many closely related reactions with a common biological function. Proteins can also bind to, or even be integrated into, cell membranes. The ability of binding partners to induce conformational changes in proteins allows the construction of enormously complex signaling networks.[31]:830–49 As interactions between proteins are reversible, and depend heavily on the availability of different groups of partner proteins to form aggregates that are capable to carry out discrete sets of function, study of the interactions between specific proteins is a key to understand important aspects of cellular function, and ultimately the properties that distinguish particular cell types.[46][47]

Enzymes

The best-known role of proteins in the cell is as enzymes, which catalyse chemical reactions. Enzymes are usually highly specific and accelerate only one or a few chemical reactions. Enzymes carry out most of the reactions involved in metabolism, as well as manipulating DNA in processes such as DNA replication, DNA repair, and transcription. Some enzymes act on other proteins to add or remove chemical groups in a process known as posttranslational modification. About 4,000 reactions are known to be catalysed by enzymes.[48] The rate acceleration conferred by enzymatic catalysis is often enormous—as much as 10^{17}-fold increase in rate over the uncatalysed reaction in the case of orotate decarboxylase (78 million years without the enzyme, 18 milliseconds with the enzyme).[49]

The molecules bound and acted upon by enzymes are called substrates. Although enzymes can consist of hundreds of amino acids, it is usually only a small fraction of the residues that come in contact with the substrate, and an even smaller fraction—three to four residues on average—that are directly involved in catalysis.[50] The region of the enzyme that binds the substrate and contains the catalytic residues is known as the active site.

Dirigent proteins are members of a class of proteins that dictate the stereochemistry of a compound synthesized by other enzymes.[51]

Cell signaling and ligand binding

Ribbon diagram of a mouse antibody against cholera that binds a carbohydrate antigen

Many proteins are involved in the process of cell signaling and signal transduction. Some proteins, such as insulin, are extracellular proteins that transmit a signal from the cell in which they were synthesized to other cells in distant tissues. Others are membrane proteins that act as receptors whose main function is to bind a signaling molecule and induce a biochemical response in the cell. Many receptors have a binding site exposed on the cell surface and an effector domain within the cell, which may have enzymatic activity or may undergo a conformational change detected by other proteins within the cell.[30]:251–81

Antibodies are protein components of an adaptive immune system whose main function is to bind antigens, or foreign substances in the body, and target them for destruction. Antibodies can be secreted into the extracellular environment or anchored in the membranes of specialized B cells known as plasma cells. Whereas enzymes are limited in their binding affinity for their substrates by the necessity of conducting their reaction, antibodies have no such constraints. An antibody's binding affinity to its target is extraordinarily high.[31]:275–50

Many ligand transport proteins bind particular small biomolecules and transport them to other locations in the body of a multicellular organism. These proteins must have a high binding affinity when their ligand is present in high concentrations, but must also release the ligand when it is present at low concentrations in the target tissues. The canonical example of a ligand-binding protein is haemoglobin, which transports oxygen from the lungs to other organs and tissues in all vertebrates and has close homologs in every biological kingdom.[31]:222–29 Lectins are sugar-binding proteins which are highly specific for their sugar moieties. Lectins typically play a role in biological recognition phenomena involving cells and proteins.[52] Receptors and hormones are highly specific binding proteins.

Transmembrane proteins can also serve as ligand transport proteins that alter the permeability of the cell membrane to small molecules and ions. The membrane alone has a hydrophobic core through which polar or charged molecules cannot diffuse. Membrane proteins contain internal channels that allow such molecules to enter and exit the cell. Many ion channel proteins are specialized to select for only a particular ion; for example, potassium and sodium channels often discriminate for only one of the two ions.[30]:232–34

Structural proteins

Structural proteins confer stiffness and rigidity to otherwise-fluid biological components. Most structural proteins are fibrous proteins; for example, collagen and elastin are critical components of connective tissue such as cartilage, and keratin is found in hard or filamentous structures such as hair, nails, feathers, hooves, and some animal shells.[31]:178–81 Some globular proteins can also play structural functions, for example, actin and tubulin are globular and soluble as monomers, but polymerize to form long, stiff fibers that make up the cytoskeleton, which allows the cell to maintain its shape and size.

Other proteins that serve structural functions are motor proteins such as myosin, kinesin, and dynein, which are capable of generating mechanical forces. These proteins are crucial for cellular motility of single celled organisms and the sperm of many multicellular organisms which

reproduce sexually. They also generate the forces exerted by contracting muscles[31]:258–64,272 and play essential roles in intracellular transport.

Protein evolution

A key question in molecular biology is how proteins evolve, i.e. how can mutations (or rather changes in amino acid sequence) lead to new structures and functions? Most amino acids in a protein can be changed without disrupting activity or function, as can be seen from numerous homologous proteins across species (as collected in specialized databases for protein families, e.g. PFAM).[53] In order to prevent dramatic consequences of mutations, a gene may be duplicated before it can mutate freely. However, this can also lead to complete loss of gene function and thus pseudo-genes.[54] More commonly, single amino acid changes have limited consequences although some can change protein function substantially, especially in enzymes. For instance, many enzymes can change their substrate specificity by one or a few mutations.[55] Changes in substrate specificity are facilitated by *substrate promiscuity*, i.e. the ability of many enzymes to bind and process multiple substrates. When mutations occur, the specificity of an enzyme can increase (or decrease) and thus its enzymatic activity.[55] Thus, bacteria (or other organisms) can adapt to different food sources, including unnatural substrates such as plastic.[56]

Methods of study

The activities and structures of proteins may be examined *in vitro*, *in vivo*, and *in silico*. ***In vitro*** studies of purified proteins in controlled environments are useful for learning how a protein carries out its function: for example, enzyme kinetics studies explore the chemical mechanism of an enzyme's catalytic activity and its relative affinity for various possible substrate molecules. By contrast, ***in vivo*** experiments can provide information about the physiological role of a protein in the context of a cell or even a whole organism. ***In silico*** studies use computational methods to study proteins.

Protein purification

To perform *in vitro* analysis, a protein must be purified away from other cellular components. This process usually begins with cell lysis, in which a cell's membrane is disrupted and its internal contents released into a solution known as a crude lysate. The resulting mixture can be purified using ultracentrifugation, which fractionates the various cellular components into

fractions containing soluble proteins; membrane lipids and proteins; cellular organelles, and nucleic acids. Precipitation by a method known as salting out can concentrate the proteins from this lysate. Various types of chromatography are then used to isolate the protein or proteins of interest based on properties such as molecular weight, net charge and binding affinity.[27]:21–24 The level of purification can be monitored using various types of gel electrophoresis if the desired protein's molecular weight and isoelectric point are known, by spectroscopy if the protein has distinguishable spectroscopic features, or by enzyme assays if the protein has enzymatic activity. Additionally, proteins can be isolated according to their charge using electrofocusing.[57]

For natural proteins, a series of purification steps may be necessary to obtain protein sufficiently pure for laboratory applications. To simplify this process, genetic engineering is often used to add chemical features to proteins that make them easier to purify without affecting their structure or activity. Here, a "tag" consisting of a specific amino acid sequence, often a series of histidine residues (a "His-tag"), is attached to one terminus of the protein. As a result, when the lysate is passed over a chromatography column containing nickel, the histidine residues ligate the nickel and attach to the column while the untagged components of the lysate pass unimpeded. A number of different tags have been developed to help researchers purify specific proteins from complex mixtures.[58]

Cellular localization

Proteins in different cellular compartments and structures tagged with green fluorescent protein (here, white)

The study of proteins *in vivo* is often concerned with the synthesis and localization of the protein within the cell. Although many intracellular proteins are synthesized in the cytoplasm and membrane-bound or secreted proteins in the endoplasmic reticulum, the specifics of how proteins are targeted to specific organelles or cellular structures is often unclear. A useful technique for assessing cellular localization uses genetic engineering to express in a cell a fusion protein or chimera consisting of the natural protein of interest linked to a "reporter" such as green fluorescent protein (GFP).[59] The fused protein's position within the cell can then be cleanly and efficiently visualized using microscopy,[60] as shown in the figure opposite.

Other methods for elucidating the cellular location of proteins requires the use of known compartmental markers for regions such as the ER, the Golgi, lysosomes or vacuoles, mitochondria, chloroplasts, plasma membrane, etc. With the use of fluorescently tagged versions of these markers or of antibodies to known markers, it becomes much simpler to identify the localization of a protein of interest. For example, indirect immunofluorescence will allow for fluorescence colocalization and demonstration of location. Fluorescent dyes are used to label cellular compartments for a similar purpose.[61]

Other possibilities exist, as well. For example, immunohistochemistry usually uses an antibody to one or more proteins of interest that are conjugated to enzymes yielding either luminescent or

chromogenic signals that can be compared between samples, allowing for localization information. Another applicable technique is cofractionation in sucrose (or other material) gradients using isopycnic centrifugation.[62] While this technique does not prove colocalization of a compartment of known density and the protein of interest, it does increase the likelihood, and is more amenable to large-scale studies.

Finally, the gold-standard method of cellular localization is immunoelectron microscopy. This technique also uses an antibody to the protein of interest, along with classical electron microscopy techniques. The sample is prepared for normal electron microscopic examination, and then treated with an antibody to the protein of interest that is conjugated to an extremely electro-dense material, usually gold. This allows for the localization of both ultrastructural details as well as the protein of interest.[63]

Through another genetic engineering application known as site-directed mutagenesis, researchers can alter the protein sequence and hence its structure, cellular localization, and susceptibility to regulation. This technique even allows the incorporation of unnatural amino acids into proteins, using modified tRNAs,[64] and may allow the rational design of new proteins with novel properties.[65]

Proteomics

The total complement of proteins present at a time in a cell or cell type is known as its proteome, and the study of such large-scale data sets defines the field of proteomics, named by analogy to the related field of genomics. Key experimental techniques in proteomics include 2D electrophoresis,[66] which allows the separation of many proteins, mass spectrometry,[67] which allows rapid high-throughput identification of proteins and sequencing of peptides (most often after in-gel digestion), protein microarrays, which allow the detection of the relative levels of the various proteins present in a cell, and two-hybrid screening, which allows the systematic exploration of protein–protein interactions.[68] The total complement of biologically possible such interactions is known as the interactome.[69] A systematic attempt to determine the structures of proteins representing every possible fold is known as structural genomics.[70]

Structure determination

Discovering the tertiary structure of a protein, or the quaternary structure of its complexes, can provide important clues about how the protein performs its function and how it can be affected,

i.e. in drug design. As proteins are too small to be seen under a light microscope, other methods have to be employed to determine their structure. Common experimental methods include X-ray crystallography and NMR spectroscopy, both of which can produce structural information at atomic resolution. However, NMR experiments are able to provide information from which a subset of distances between pairs of atoms can be estimated, and the final possible conformations for a protein are determined by solving a distance geometry problem. Dual polarisation interferometry is a quantitative analytical method for measuring the overall protein conformation and conformational changes due to interactions or other stimulus. Circular dichroism is another laboratory technique for determining internal β-sheet / α-helical composition of proteins. Cryoelectron microscopy is used to produce lower-resolution structural information about very large protein complexes, including assembled viruses;[30]:340–41 a variant known as electron crystallography can also produce high-resolution information in some cases, especially for two-dimensional crystals of membrane proteins.[71] Solved structures are usually deposited in the Protein Data Bank (PDB), a freely available resource from which structural data about thousands of proteins can be obtained in the form of Cartesian coordinates for each atom in the protein.[72]

Many more gene sequences are known than protein structures. Further, the set of solved structures is biased toward proteins that can be easily subjected to the conditions required in X-ray crystallography, one of the major structure determination methods. In particular, globular proteins are comparatively easy to crystallize in preparation for X-ray crystallography. Membrane proteins and large protein complexes, by contrast, are difficult to crystallize and are underrepresented in the PDB.[73] Structural genomics initiatives have attempted to remedy these deficiencies by systematically solving representative structures of major fold classes. Protein structure prediction methods attempt to provide a means of generating a plausible structure for proteins whose structures have not been experimentally determined.[74]

Structure prediction

Constituent amino-acids can be analyzed to predict secondary, tertiary and quaternary protein structure, in this case hemoglobin containing heme units

Complementary to the field of structural genomics, *protein structure prediction* develops efficient mathematical models of proteins to computationally predict the molecular formations in theory, instead of detecting structures with laboratory observation.[75] The most successful type of structure prediction, known as homology modeling, relies on the existence of a "template" structure with sequence similarity to the protein being modeled; structural genomics' goal is to provide sufficient representation in solved structures to model most of those that remain.[76] Although producing accurate models remains a challenge when only distantly related template structures are available, it has been suggested that sequence alignment is the bottleneck in this process, as quite accurate models can be produced if a "perfect" sequence alignment is known.[77] Many structure prediction methods have served to inform the emerging field of protein engineering, in which novel protein folds have already been designed.[78] Also proteins (in eukaryotes ~33%) contain large unstructured but biologically functional segments and can be classified as intrinsically disordered proteins.[79] Predicting and analysing protein disorder is, therefore, an important part of protein structure characterisation.[80]

Bioinformatics

A vast array of computational methods have been developed to analyze the structure, function and evolution of proteins. The development of such tools has been driven by the large amount of genomic and proteomic data available for a variety of organisms, including the human genome. It is simply impossible to study all proteins experimentally, hence only a few are subjected to

laboratory experiments while computational tools are used to extrapolate to similar proteins. Such homologous proteins can be efficiently identified in distantly related organisms by sequence alignment. Genome and gene sequences can be searched by a variety of tools for certain properties. Sequence profiling tools can find restriction enzyme sites, open reading frames in nucleotide sequences, and predict secondary structures. Phylogenetic trees can be constructed and evolutionary hypotheses developed using special software like ClustalW regarding the ancestry of modern organisms and the genes they express. The field of bioinformatics is now indispensable for the analysis of genes and proteins.

In silico simulation of dynamical processes

A more complex computational problem is the prediction of intermolecular interactions, such as in molecular docking,[81] protein folding, protein–protein interaction and chemical reactivity. Mathematical models to simulate these dynamical processes involve molecular mechanics, in particular, molecular dynamics. In this regard, *in silico* simulations discovered the folding of small α-helical protein domains such as the villin headpiece,[82] the HIV accessory protein[83] and hybrid methods combining standard molecular dynamics with quantum mechanical mathematics have explored the electronic states of rhodopsins.[84]

Beyond classical molecular dynamics, quantum dynamics methods allow the simulation of proteins in atomistic detail with an accurate description of quantum mechanical effects. Examples include the multi-layer multi-configuration time-dependent Hartree (MCTDH) method and the hierarchical equations of motion (HEOM) approach, which have been applied to plant cryptochromes[85] and bacteria light-harvesting complexes,[86] respectively. Both quantum and classical mechanical simulations of biological-scale systems are extremely computationally demanding, so distributed computing initiatives (for example, the Folding@home project[87]) facilitate the molecular modeling by exploiting advances in GPU parallel processing and Monte Carlo techniques.

Chemical analysis

The total nitrogen content of organic matter is mainly formed by the amino groups in proteins. The Total Kjeldahl Nitrogen (TKN) is a measure of nitrogen widely used in the analysis of (waste) water, soil, food, feed and organic matter in general. As the name suggests, the Kjeldahl method is applied. More sensitive methods are available.[88][89]

Nutrition

Most microorganisms and plants can biosynthesize all 20 standard amino acids, while animals (including humans) must obtain some of the amino acids from the diet.[44] The amino acids that an organism cannot synthesize on its own are referred to as essential amino acids. Key enzymes that synthesize certain amino acids are not present in animals—such as aspartokinase, which catalyses the first step in the synthesis of lysine, methionine, and threonine from aspartate. If amino acids are present in the environment, microorganisms can conserve energy by taking up the amino acids from their surroundings and downregulating their biosynthetic pathways.

In animals, amino acids are obtained through the consumption of foods containing protein. Ingested proteins are then broken down into amino acids through digestion, which typically involves denaturation of the protein through exposure to acid and hydrolysis by enzymes called proteases. Some ingested amino acids are used for protein biosynthesis, while others are converted to glucose through gluconeogenesis, or fed into the citric acid cycle. This use of protein as a fuel is particularly important under starvation conditions as it allows the body's own proteins to be used to support life, particularly those found in muscle.[90]

In animals such as dogs and cats, protein maintains the health and quality of the skin by promoting hair follicle growth and keratinization, and thus reducing the likelihood of skin problems producing malodours.[91] Poor-quality proteins also have a role regarding gastrointestinal health, increasing the potential for flatulence and odorous compounds in dogs because when proteins reach the colon in an undigested state, they are fermented producing hydrogen sulfide gas, indole, and skatole.[92] Dogs and cats digest animal proteins better than those from plants, but products of low-quality animal origin are poorly digested, including skin, feathers, and connective tissue.[92]

See also

- Deproteination
- DNA-binding protein
- Macromolecule
- Index of protein-related articles
- Intein

- List of proteins
- Proteopathy
- Proteopedia
- Proteolysis
- Protein sequence space
- Protein superfamily

References

1. Thomas Burr Osborne (1909): The Vegetable Proteins (https://archive.org/details/vegetableprotein00osbouoft) Archived (https://web.archive.org/web/20160322224726/https://archive.org/details/vegetableprotein00osbouoft) 2016-03-22 at the Wayback Machine, History pp 1 to 6, from archive.org
2. Mulder GJ (1838). "Sur la composition de quelques substances animales" (https://archive.org/stream/bulletindesscien00leyd#page/104/mode/2up) . *Bulletin des Sciences Physiques et Naturelles en Néerlande*: 104.
3. Harold H (1951). "Origin of the Word 'Protein.' " (https://doi.org/10.1038%2F168244a0) . *Nature*. **168** (4267): 244. Bibcode:1951Natur.168..244H (https://ui.adsabs.harvard.edu/abs/1951Natur.168..244H) . doi:10.1038/168244a0 (https://doi.org/10.1038%2F168244a0) . PMID 14875059 (https://pubmed.ncbi.nlm.nih.gov/14875059) . S2CID 4271525 (https://api.semanticscholar.org/CorpusID:4271525) .
4. Perrett D (August 2007). "From 'protein' to the beginnings of clinical proteomics". *Proteomics: Clinical Applications*. **1** (8): 720–38. doi:10.1002/prca.200700525 (https://doi.org/10.1002%2Fprca.200700525) . PMID 21136729 (https://pubmed.ncbi.nlm.nih.gov/21136729) . S2CID 32843102 (https://api.semanticscholar.org/CorpusID:32843102) .
5. *New Oxford Dictionary of English*
6. Reynolds JA, Tanford C (2003). *Nature's Robots: A History of Proteins (Oxford Paperbacks)*. New York, New York: Oxford University Press. p. 15. ISBN 978-0-19-860694-9.
7. Reynolds and Tanford (2003).
8. Bischoff TL, Voit C (1860). *Die Gesetze der Ernaehrung des Pflanzenfressers durch neue Untersuchungen festgestellt* (in German). Leipzig, Heidelberg.
9. "Hofmeister, Franz" (http://www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/hofmeister-franz) . encyclopedia.com. Archived (https://web.archive.org/web/20170405073423/http://www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/hofmeister-franz) from the original on 5 April 2017. Retrieved 4 April 2017.

10. "Protein, section: Classification of protein" (https://www.britannica.com/science/protein/Conformation-of-proteins-in-interfaces#ref593795) . britannica.com. Archived (https://web.archive.org/web/20170404225132/https://www.britannica.com/science/protein/Conformation-of-proteins-in-interfaces#ref593795) from the original on 4 April 2017. Retrieved 4 April 2017.

11. Sumner, James B. (August 1926). "The Isolation and Crystallization of the Enzyme Urease" (https://doi.org/10.1016%2FS0021-9258%2818%2984560-4) . *Journal of Biological Chemistry*. **69** (2): 435–441. doi:10.1016/S0021-9258(18)84560-4 (https://doi.org/10.1016%2FS0021-9258%2818%2984560-4) .

12. Pauling, Linus; Corey, Robert B. (May 1951). "Atomic Coordinates and Structure Factors for Two Helical Configurations of Polypeptide Chains" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063348) . *Proceedings of the National Academy of Sciences*. **37** (5): 235–240. Bibcode:1951PNAS...37..235P (https://ui.adsabs.harvard.edu/abs/1951PNAS...37..235P) . doi:10.1073/pnas.37.5.235 (https://doi.org/10.1073%2Fpnas.37.5.235) . PMC 1063348 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063348) . PMID 14834145 (https://pubmed.ncbi.nlm.nih.gov/14834145) .

13. Kauzmann W (May 1956). "Structural factors in protein denaturation". *Journal of Cellular Physiology*. **47** (Suppl 1): 113–31. doi:10.1002/jcp.1030470410 (https://doi.org/10.1002%2Fjcp.1030470410) . PMID 13332017 (https://pubmed.ncbi.nlm.nih.gov/13332017) .

14. Kauzmann W (1959). "Some factors in the interpretation of protein denaturation". *Advances in Protein Chemistry Volume 14*. Vol. 14. pp. 1–63. doi:10.1016/S0065-3233(08)60608-7 (https://doi.org/10.1016%2FS0065-3233%2808%2960608-7) . ISBN 978-0-12-034214-3. PMID 14404936 (https://pubmed.ncbi.nlm.nih.gov/14404936) .

15. Kalman SM, Linderstrøm-Lang K, Ottesen M, Richards FM (February 1955). "Degradation of ribonuclease by subtilisin". *Biochimica et Biophysica Acta*. **16** (2): 297–99. doi:10.1016/0006-3002(55)90224-9 (https://doi.org/10.1016%2F0006-3002%2855%2990224-9) . PMID 14363272 (https://pubmed.ncbi.nlm.nih.gov/14363272) .

16. Sanger F (1949). "The terminal peptides of insulin" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1275055) . *The Biochemical Journal*. **45** (5): 563–74. doi:10.1042/bj0450563 (https://doi.org/10.1042%2Fbj0450563) . PMC 1275055 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1275055) . PMID 15396627 (https://pubmed.ncbi.nlm.nih.gov/15396627) .

17. Sanger F. (1958), *Nobel lecture: The chemistry of insulin* (http://nobelprize.org/nobel_prizes/chemistry/laureates/1958/sanger-lecture.pdf) (PDF), Nobelprize.org, archived (https://web.archive.org/web/20130319022148/http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1958/sanger-lecture.pdf) (PDF) from the original on 2013-03-19, retrieved 2016-02-09

18. Stoddart, Charlotte (1 March 2022). "Structural biology: How proteins got their close-up" (https://doi.org/10.1146%2Fknowable-022822-1) . *Knowable Magazine*. doi:10.1146/knowable-022822-1 (https://doi.org/10.1146%2Fknowable-022822-1) .

19. Muirhead H, Perutz MF (August 1963). "Structure of hemoglobin. A three-dimensional fourier synthesis of reduced human hemoglobin at 5.5 Å resolution". *Nature*. **199** (4894): 633–38. Bibcode:1963Natur.199..633M (https://ui.adsabs.harvard.edu/abs/1963Natur.199..633M) . doi: 10.1038/199633a0 (https://doi.org/10.1038%2F199633a0) . PMID 14074546 (https://pubmed.ncbi.nlm.nih.gov/14074546) . S2CID 4257461 (https://api.semanticscholar.org/CorpusID:4257461) .

20. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (March 1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis". *Nature*. **181** (4610): 662–66. Bibcode:1958Natur.181..662K (https://ui.adsabs.harvard.edu/abs/1958Natur.181..662K) . doi: 10.1038/181662a0 (https://doi.org/10.1038%2F181662a0) . PMID 13517261 (https://pubmed.ncbi.nlm.nih.gov/13517261) . S2CID 4162786 (https://api.semanticscholar.org/CorpusID:4162786) .

21. Zhou ZH (April 2008). "Towards atomic resolution structural determination by single-particle cryo-electron microscopy" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2714865) . *Current Opinion in Structural Biology*. **18** (2): 218–28. doi:10.1016/j.sbi.2008.03.004 (https://doi.org/10.1016%2Fj.sbi.2008.03.004) . PMC 2714865 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2714865) . PMID 18403197 (https://pubmed.ncbi.nlm.nih.gov/18403197) .

22. Keskin O, Tuncbag N, Gursoy A (April 2008). "Characterization and prediction of protein interfaces to infer protein-protein interaction networks". *Current Pharmaceutical Biotechnology*. **9** (2): 67–76. doi: 10.2174/138920108783955191 (https://doi.org/10.2174%2F138920108783955191) . hdl: 11511/32640 (https://hdl.handle.net/11511%2F32640) . PMID 18393863 (https://pubmed.ncbi.nlm.nih.gov/18393863) .

23. "RCSB Protein Data Bank" (https://web.archive.org/web/20150418160606/http://www.rcsb.org/pdb/home/home.do) . Archived from the original (http://www.rcsb.org/pdb/home/home.do) on 2015-04-18. Retrieved 2017-01-19.

24. Ekman, Diana; Björklund, Åsa K.; Frey-Skött, Johannes; Elofsson, Arne (April 2005). "Multi-domain Proteins in the Three Kingdoms of Life: Orphan Domains and Other Unassigned Regions". *Journal of Molecular Biology*. **348** (1): 231–243. doi:10.1016/j.jmb.2005.02.007 (https://doi.org/10.1016%2Fj.jmb.2005.02.007) . PMID 15808866 (https://pubmed.ncbi.nlm.nih.gov/15808866) .

25. Nelson DL, Cox MM (2005). *Lehninger's Principles of Biochemistry* (4th ed.). New York, New York: W. H. Freeman and Company.

26. Gutteridge A, Thornton JM (November 2005). "Understanding nature's catalytic toolkit". *Trends in Biochemical Sciences*. **30** (11): 622–29. doi:10.1016/j.tibs.2005.09.006 (https://doi.org/10.1016%2Fj.tibs.2005.09.006) . PMID 16214343 (https://pubmed.ncbi.nlm.nih.gov/16214343) .

27. Murray RF, Harper HW, Granner DK, Mayes PA, Rodwell VW (2006). *Harper's Illustrated Biochemistry*. New York: Lange Medical Books/McGraw-Hill. ISBN 978-0-07-146197-9.

28. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J (2004). *Molecular Cell Biology* (5th ed.). New York, New York: WH Freeman and Company.

29. Ardejani MS, Powers ET, Kelly JW (August 2017). "Using Cooperatively Folded Peptides To Measure Interaction Energies and Conformational Propensities" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5584629) . *Accounts of Chemical Research*. **50** (8): 1875–1882. doi:10.1021/acs.accounts.7b00195 (https://doi.org/10.1021%2Facs.accounts.7b00195) . PMC 5584629 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5584629) . PMID 28723063 (https://pubmed.ncbi.nlm.nih.gov/28723063) .

30. Branden C, Tooze J (1999). *Introduction to Protein Structure*. New York: Garland Pub. ISBN 978-0-8153-2305-1.

31. Van Holde KE, Mathews CK (1996). *Biochemistry* (https://archive.org/details/biochemistry00math) . Menlo Park, California: Benjamin/Cummings Pub. Co., Inc. ISBN 978-0-8053-3931-4.

32. Milo R (December 2013). "What is the total number of protein molecules per cell volume? A call to rethink some published values" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3910158) . *BioEssays*. **35** (12): 1050–55. doi:10.1002/bies.201300066 (https://doi.org/10.1002%2Fbies.201300066) . PMC 3910158 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3910158) . PMID 24114984 (https://pubmed.ncbi.nlm.nih.gov/24114984) .

33. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R (November 2011). "The quantitative proteome of a human cell line" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261713) . *Molecular Systems Biology*. **7**: 549. doi:10.1038/msb.2011.82 (https://doi.org/10.1038%2Fmsb.2011.82) . PMC 3261713 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261713) . PMID 22068332 (https://pubmed.ncbi.nlm.nih.gov/22068332) .

34. Wu L, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, Tang H, Snyder M (July 2013). "Variation and genetic control of protein abundance in humans" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3789121) . *Nature*. **499** (7456): 79–82. Bibcode:2013Natur.499...79W (https://ui.adsabs.harvard.edu/abs/2013Natur.499...79W) . doi:10.1038/nature12223 (https://doi.org/10.1038%2Fnature12223) . PMC 3789121 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3789121) . PMID 23676674 (https://pubmed.ncbi.nlm.nih.gov/23676674) .

35. Dobson CM (2000). "The nature and significance of protein folding". In Pain RH (ed.). *Mechanisms of Protein Folding*. Oxford, Oxfordshire: Oxford University Press. pp. 1–28. ISBN 978-0-19-963789-8.

36. Kozlowski LP (January 2017). "Proteome-pI: proteome isoelectric point database" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210655) . *Nucleic Acids Research*. **45** (D1): D1112–D1116. doi:10.1093/nar/gkw978 (https://doi.org/10.1093%2Fnar%2Fgkw978) . PMC 5210655 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5210655) . PMID 27789699 (https://pubmed.ncbi.nlm.nih.gov/27789699) .

37. Fulton AB, Isaacs WB (April 1991). "Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis". *BioEssays*. **13** (4): 157–61. doi:10.1002/bies.950130403 (https://doi.org/10.1002%2Fbies.950130403) . PMID 1859393 (https://pubmed.ncbi.nlm.nih.gov/1859393) . S2CID 20237314 (https://api.semanticscholar.org/CorpusID:20237314) .

38. Bruckdorfer T, Marder O, Albericio F (February 2004). "From production of peptides in milligram amounts for research to multi-tons quantities for drugs of the future". *Current Pharmaceutical Biotechnology*. **5** (1): 29–43. doi:10.2174/1389201043489620 (https://doi.org/10.2174%2F1389201043489620) . PMID 14965208 (https://pubmed.ncbi.nlm.nih.gov/14965208) .

39. Schwarzer D, Cole PA (December 2005). "Protein semisynthesis and expressed protein ligation: chasing a protein's tail". *Current Opinion in Chemical Biology*. **9** (6): 561–69. doi:10.1016/j.cbpa.2005.09.018 (https://doi.org/10.1016%2Fj.cbpa.2005.09.018) . PMID 16226484 (https://pubmed.ncbi.nlm.nih.gov/16226484) .

40. Kent SB (February 2009). "Total chemical synthesis of proteins". *Chemical Society Reviews*. **38** (2): 338–51. doi:10.1039/b700141j (https://doi.org/10.1039%2Fb700141j) . PMID 19169452 (https://pubmed.ncbi.nlm.nih.gov/19169452) . S2CID 5432012 (https://api.semanticscholar.org/CorpusID:5432012) .

41. Fernández A, Scott R (September 2003). "Dehydron: a structurally encoded signal for protein interaction" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1303363) . *Biophysical Journal*. **85** (3): 1914–28. Bibcode:2003BpJ....85.1914F (https://ui.adsabs.harvard.edu/abs/2003BpJ....85.1914F) . doi:10.1016/S0006-3495(03)74619-0 (https://doi.org/10.1016%2FS0006-3495%2803%2974619-0) . PMC 1303363 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1303363) . PMID 12944304 (https://pubmed.ncbi.nlm.nih.gov/12944304) .

42. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ (January 2012). "Attributes of short linear motifs". *Molecular BioSystems*. **8** (1): 268–81. doi:10.1039/c1mb05231d (https://doi.org/10.1039%2Fc1mb05231d) . PMID 21909575 (https://pubmed.ncbi.nlm.nih.gov/21909575) .

43. Scalvini, Barbara; Sheikhhassani, Vahid; Woodard, Jaie; Aupič, Jana; Dame, Remus T.; Jerala, Roman; Mashaghi, Alireza (July 2020). "Topology of Folded Molecular Chains: From Single Biomolecules to Engineered Origami". *Trends in Chemistry*. **2** (7): 609–622. doi:10.1016/j.trechm.2020.04.009 (https://doi.org/10.1016%2Fj.trechm.2020.04.009) . hdl:1887/3245505 (https://hdl.handle.net/1887%2F3245505) . S2CID 218957613 (https://api.semanticscholar.org/CorpusID:218957613) .

44. Voet D, Voet JG. (2004). *Biochemistry* Vol 1 3rd ed. Wiley: Hoboken, NJ.

45. Sankaranarayanan R, Moras D (2001). "The fidelity of the translation of the genetic code" (https://doi.org/10.18388%2Fabp.2001_3918) . *Acta Biochimica Polonica*. **48** (2): 323–35. doi:10.18388/abp.2001_3918 (https://doi.org/10.18388%2Fabp.2001_3918) . PMID 11732604 (https://pubmed.ncbi.nlm.nih.gov/11732604) .

46. Copland JA, Sheffield-Moore M, Koldzic-Zivanovic N, Gentry S, Lamprou G, Tzortzatou-Stathopoulou F, Zoumpourlis V, Urban RJ, Vlahopoulos SA (June 2009). "Sex steroid receptors in skeletal differentiation and epithelial neoplasia: is tissue-specific intervention possible?". *BioEssays*. **31** (6): 629–41. doi: 10.1002/bies.200800138 (https://doi.org/10.1002%2Fbies.200800138) . PMID 19382224 (https://pubmed.ncbi.nlm.nih.gov/19382224) . S2CID 205469320 (https://api.semanticscholar.org/CorpusID:205469320) .

47. Samarin S, Nusrat A (January 2009). "Regulation of epithelial apical junctional complex by Rho family GTPases" (https://doi.org/10.2741%2F3298) . *Frontiers in Bioscience*. **14** (14): 1129–42. doi: 10.2741/3298 (https://doi.org/10.2741%2F3298) . PMID 19273120 (https://pubmed.ncbi.nlm.nih.gov/19273120) .

48. Bairoch, A. (2000). "The ENZYME database in 2000" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102465) . *Nucleic Acids Research*. **28** (1): 304–305. doi:10.1093/nar/28.1.304 (https://doi.org/10.1093%2Fnar%2F28.1.304) . PMC 102465 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102465) . PMID 10592255 (https://pubmed.ncbi.nlm.nih.gov/10592255) .

49. Radzicka A, Wolfenden R (January 1995). "A proficient enzyme". *Science*. **267** (5194): 90–3. Bibcode:1995Sci...267...90R (https://ui.adsabs.harvard.edu/abs/1995Sci...267...90R) . doi:10.1126/science.7809611 (https://doi.org/10.1126%2Fscience.7809611) . PMID 7809611 (https://pubmed.ncbi.nlm.nih.gov/7809611) .

50. EBI External Services (2010-01-20). "The Catalytic Site Atlas at The European Bioinformatics Institute" (http://www.ebi.ac.uk/thornton-srv/databases/CSA/) . Ebi.ac.uk. Archived (https://web.archive.org/web/20130803032349/http://www.ebi.ac.uk/thornton-srv/databases/CSA/) from the original on 2013-08-03. Retrieved 2011-01-16.

51. Pickel B, Schaller A (October 2013). "Dirigent proteins: molecular characteristics and potential biotechnological applications". *Applied Microbiology and Biotechnology*. **97** (19): 8427–38. doi: 10.1007/s00253-013-5167-4 (https://doi.org/10.1007%2Fs00253-013-5167-4) . PMID 23989917 (https://pubmed.ncbi.nlm.nih.gov/23989917) . S2CID 1896003 (https://api.semanticscholar.org/CorpusID:1896003) .

52. Rüdiger H, Siebert HC, Solís D, Jiménez-Barbero J, Romero A, von der Lieth CW, Diaz-Mariño T, Gabius HJ (April 2000). "Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets". *Current Medicinal Chemistry*. **7** (4): 389–416. doi: 10.2174/0929867003375164 (https://doi.org/10.2174%2F0929867003375164) . PMID 10702616 (https://pubmed.ncbi.nlm.nih.gov/10702616) .

53. Mulder NJ (2007-09-28). "Protein Family Databases". *eLS*. Chichester, UK: John Wiley & Sons, Ltd. pp. a0003058.pub2. doi:10.1002/9780470015902.a0003058.pub2 (https://doi.org/10.1002%2F9780470015902.a0003058.pub2) . ISBN 978-0-470-01617-6.

54. Sisu C, Pei B, Leng J, Frankish A, Zhang Y, Balasubramanian S, et al. (September 2014). "Comparative analysis of pseudogenes across three phyla" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4169933) . *Proceedings of the National Academy of Sciences of the United States of America*. **111** (37): 13361–6. Bibcode:2014PNAS..11113361S (https://ui.adsabs.harvard.edu/abs/2014PNAS..111133361S) . doi:10.1073/pnas.1407293111 (https://doi.org/10.1073%2Fpnas.1407293111) . PMC 4169933 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4169933) . PMID 25157146 (https://pubmed.ncbi.nlm.nih.gov/25157146) .

55. Guzmán GI, Sandberg TE, LaCroix RA, Nyerges Á, Papp H, de Raad M, et al. (April 2019). "Enzyme promiscuity shapes adaptation to novel growth substrates" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6452873) . *Molecular Systems Biology*. **15** (4): e8462. doi:10.15252/msb.20188462 (https://doi.org/10.15252%2Fmsb.20188462) . PMC 6452873 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6452873) . PMID 30962359 (https://pubmed.ncbi.nlm.nih.gov/30962359) .

56. Roohi, Bano K, Kuddus M, Zaheer MR, Zia Q, Khan MF, Ashraf GM, Gupta A, Aliev G (2017). "Microbial Enzymatic Degradation of Biodegradable Plastics". *Current Pharmaceutical Biotechnology*. **18** (5): 429–440. doi:10.2174/1389201018666170523165742 (https://doi.org/10.2174%2F1389201018666170523165742) . PMID 28545359 (https://pubmed.ncbi.nlm.nih.gov/28545359) .

57. Hey, Julie; Posch, Anton; Cohen, Andrew; Liu, Ning; Harbers, Adrianna (2008). "Fractionation of Complex Protein Mixtures by Liquid-Phase Isoelectric Focusing". *2D PAGE: Sample Preparation and Fractionation*. Methods in Molecular Biology. Vol. 424. pp. 225–239. doi:10.1007/978-1-60327-064-9_19 (https://doi.org/10.1007%2F978-1-60327-064-9_19) . ISBN 978-1-58829-722-8. PMID 18369866 (https://pubmed.ncbi.nlm.nih.gov/18369866) .

58. Terpe K (January 2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems". *Applied Microbiology and Biotechnology*. **60** (5): 523–33. doi:10.1007/s00253-002-1158-6 (https://doi.org/10.1007%2Fs00253-002-1158-6) . PMID 12536251 (https://pubmed.ncbi.nlm.nih.gov/12536251) . S2CID 206934268 (https://api.semanticscholar.org/CorpusID:206934268) .

59. Stepanenko OV, Verkhusha VV, Kuznetsova IM, Uversky VN, Turoverov KK (August 2008). "Fluorescent proteins as biomarkers and biosensors: throwing color lights on molecular and cellular processes" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904242) . *Current Protein & Peptide Science*. **9** (4): 338–69. doi:10.2174/138920308785132668 (https://doi.org/10.2174%2F138920308785132668) . PMC 2904242 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2904242) . PMID 18691124 (https://pubmed.ncbi.nlm.nih.gov/18691124) .

60. Yuste R (December 2005). "Fluorescence microscopy today". *Nature Methods*. **2** (12): 902–4. doi:10.1038/nmeth1205-902 (https://doi.org/10.1038%2Fnmeth1205-902) . PMID 16299474 (https://pubmed.ncbi.nlm.nih.gov/16299474) . S2CID 205418407 (https://api.semanticscholar.org/CorpusID:205418407) .

61. Margolin W (January 2000). "Green fluorescent protein as a reporter for macromolecular localization in bacterial cells". *Methods*. **20** (1): 62–72. doi:10.1006/meth.1999.0906 (https://doi.org/10.1006%2Fmeth.1999.0906) . PMID 10610805 (https://pubmed.ncbi.nlm.nih.gov/10610805) .

62. Walker JH, Wilson K (2000). *Principles and Techniques of Practical Biochemistry*. Cambridge, UK: Cambridge University Press. pp. 287–89. ISBN 978-0-521-65873-7.

63. Mayhew TM, Lucocq JM (August 2008). "Developments in cell biology for quantitative immunoelectron microscopy based on thin sections: a review" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491712) . *Histochemistry and Cell Biology*. **130** (2): 299–313. doi:10.1007/s00418-008-0451-6 (https://doi.org/10.1007%2Fs00418-008-0451-6) . PMC 2491712 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491712) . PMID 18553098 (https://pubmed.ncbi.nlm.nih.gov/18553098) .

64. Hohsaka T, Sisido M (December 2002). "Incorporation of non-natural amino acids into proteins". *Current Opinion in Chemical Biology*. **6** (6): 809–15. doi:10.1016/S1367-5931(02)00376-9 (https://doi.org/10.1016%2FS1367-5931%2802%2900376-9) . PMID 12470735 (https://pubmed.ncbi.nlm.nih.gov/12470735) .

65. Cedrone F, Ménez A, Quéméneur E (August 2000). "Tailoring new enzyme functions by rational redesign". *Current Opinion in Structural Biology*. **10** (4): 405–10. doi:10.1016/S0959-440X(00)00106-8 (https://doi.org/10.1016%2FS0959-440X%2800%2900106-8) . PMID 10981626 (https://pubmed.ncbi.nlm.nih.gov/10981626) .

66. Görg A, Weiss W, Dunn MJ (December 2004). "Current two-dimensional electrophoresis technology for proteomics". *Proteomics*. **4** (12): 3665–85. doi:10.1002/pmic.200401031 (https://doi.org/10.1002%2Fpmic.200401031) . PMID 15543535 (https://pubmed.ncbi.nlm.nih.gov/15543535) . S2CID 28594824 (https://api.semanticscholar.org/CorpusID:28594824) .

67. Conrotto P, Souchelnytskyi S (September 2008). "Proteomic approaches in biological and medical sciences: principles and applications". *Experimental Oncology*. **30** (3): 171–80. PMID 18806738 (https://pubmed.ncbi.nlm.nih.gov/18806738) .

68. Koegl, M.; Uetz, P. (22 January 2008). "Improving yeast two-hybrid screening systems". *Briefings in Functional Genomics and Proteomics*. **6** (4): 302–312. doi:10.1093/bfgp/elm035 (https://doi.org/10.1093%2Fbfgp%2Felm035) . PMID 18218650 (https://pubmed.ncbi.nlm.nih.gov/18218650) .

69. Plewczyński D, Ginalski K (2009). "The interactome: predicting the protein-protein interactions in cells" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6275871) . *Cellular & Molecular Biology Letters*. **14** (1): 1–22. doi:10.2478/s11658-008-0024-7 (https://doi.org/10.2478%2Fs11658-008-0024-7) . PMC 6275871 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6275871) . PMID 18839074 (https://pubmed.ncbi.nlm.nih.gov/18839074) .

70. Zhang, Chao; Kim, Sung-Hou (February 2003). "Overview of structural genomics: from structure to function" (https://zenodo.org/record/1260238) . *Current Opinion in Chemical Biology*. **7** (1): 28–32. doi:10.1016/S1367-5931(02)00015-7 (https://doi.org/10.1016%2FS1367-5931%2802%2900015-7) . PMID 12547423 (https://pubmed.ncbi.nlm.nih.gov/12547423) .

71. Gonen T, Cheng Y, Sliz P, Hiroaki Y, Fujiyoshi Y, Harrison SC, Walz T (December 2005). "Lipid-protein interactions in double-layered two-dimensional AQP0 crystals" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1350984) . *Nature*. **438** (7068): 633–38. Bibcode:2005Natur.438..633G (https://ui.adsabs.harvard.edu/abs/2005Natur.438..633G) . doi:10.1038/nature04321 (https://doi.org/10.1038%2Fnature04321) . PMC 1350984 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1350984) . PMID 16319884 (https://pubmed.ncbi.nlm.nih.gov/16319884) .

72. Standley DM, Kinjo AR, Kinoshita K, Nakamura H (July 2008). "Protein structure databases with new web services for structural biology and biomedical research" (https://doi.org/10.1093%2Fbib%2Fbbn015) . *Briefings in Bioinformatics*. **9** (4): 276–85. doi:10.1093/bib/bbn015 (https://doi.org/10.1093%2Fbib%2Fbbn015) . PMID 18430752 (https://pubmed.ncbi.nlm.nih.gov/18430752) .

73. Walian P, Cross TA, Jap BK (2004). "Structural genomics of membrane proteins" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC395774) . *Genome Biology*. **5** (4): 215. doi:10.1186/gb-2004-5-4-215 (https://doi.org/10.1186%2Fgb-2004-5-4-215) . PMC 395774 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC395774) . PMID 15059248 (https://pubmed.ncbi.nlm.nih.gov/15059248) .

74. Sleator RD (2012). "Prediction of Protein Functions". *Functional Genomics*. Methods in Molecular Biology. Vol. 815. pp. 15–24. doi:10.1007/978-1-61779-424-7_2 (https://doi.org/10.1007%2F978-1-61779-424-7_2) . ISBN 978-1-61779-423-0. PMID 22130980 (https://pubmed.ncbi.nlm.nih.gov/22130980) .

75. Zhang Y (June 2008). "Progress and challenges in protein structure prediction" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2680823) . *Current Opinion in Structural Biology*. **18** (3): 342–48. doi:10.1016/j.sbi.2008.02.004 (https://doi.org/10.1016%2Fj.sbi.2008.02.004) . PMC 2680823 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2680823) . PMID 18436442 (https://pubmed.ncbi.nlm.nih.gov/18436442) .

76. Xiang Z (June 2006). "Advances in homology protein structure modeling" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839925) . *Current Protein & Peptide Science*. **7** (3): 217–27. doi:10.2174/138920306777452312 (https://doi.org/10.2174%2F138920306777452312) . PMC 1839925 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839925) . PMID 16787261 (https://pubmed.ncbi.nlm.nih.gov/16787261) .

77. Zhang Y, Skolnick J (January 2005). "The protein structure prediction problem could be solved using the current PDB library" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC545829) . *Proceedings of the National Academy of Sciences of the United States of America*. **102** (4): 1029–34. Bibcode:2005PNAS..102.1029Z (https://ui.adsabs.harvard.edu/abs/2005PNAS..102.1029Z) . doi:10.1073/pnas.0407152101 (https://doi.org/10.1073%2Fpnas.0407152101) . PMC 545829 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC545829) . PMID 15653774 (https://pubmed.ncbi.nlm.nih.gov/15653774) .

78. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (November 2003). "Design of a novel globular protein fold with atomic-level accuracy". *Science*. **302** (5649): 1364–68. Bibcode:2003Sci...302.1364K (https://ui.adsabs.harvard.edu/abs/2003Sci...302.1364K) . doi:10.1126/science.1089427 (https://doi.org/10.1126%2Fscience.1089427) . PMID 14631033 (https://pubmed.ncbi.nlm.nih.gov/14631033) . S2CID 1939390 (https://api.semanticscholar.org/CorpusID:1939390) .

79. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (March 2004). "Prediction and functional analysis of native disorder in proteins from the three kingdoms of life". *Journal of Molecular Biology*. **337** (3): 635–45. CiteSeerX 10.1.1.120.5605 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.120.5605) . doi:10.1016/j.jmb.2004.02.002 (https://doi.org/10.1016%2Fj.jmb.2004.02.002) . PMID 15019783 (https://pubmed.ncbi.nlm.nih.gov/15019783) .

80. Tompa, Peter; Fersht, Alan (2009). *Structure and Function of Intrinsically Disordered Proteins*. CRC Press. ISBN 978-1-4200-7893-0.

81. Ritchie DW (February 2008). "Recent progress and future directions in protein-protein docking". *Current Protein & Peptide Science*. **9** (1): 1–15. CiteSeerX 10.1.1.211.4946 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.211.4946) . doi:10.2174/138920308783565741 (https://doi.org/10.2174%2F138920308783565741) . PMID 18336319 (https://pubmed.ncbi.nlm.nih.gov/18336319) .

82. Zagrovic B, Snow CD, Shirts MR, Pande VS (November 2002). "Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing". *Journal of Molecular Biology*. **323** (5): 927–37. CiteSeerX 10.1.1.142.8664 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.142.8664) . doi:10.1016/S0022-2836(02)00997-X (https://doi.org/10.1016%2FS0022-2836%2802%2900997-X) . PMID 12417204 (https://pubmed.ncbi.nlm.nih.gov/12417204) .

83. Herges T, Wenzel W (January 2005). "In silico folding of a three helix protein and characterization of its free-energy landscape in an all-atom force field". *Physical Review Letters*. **94** (1): 018101. arXiv:physics/0310146 (https://arxiv.org/abs/physics/0310146) . Bibcode:2005PhRvL..94a8101H (https://ui.adsabs.harvard.edu/abs/2005PhRvL..94a8101H) . doi:10.1103/PhysRevLett.94.018101 (https://doi.org/10.1103%2FPhysRevLett.94.018101) . PMID 15698135 (https://pubmed.ncbi.nlm.nih.gov/15698135) . S2CID 1477100 (https://api.semanticscholar.org/CorpusID:1477100) .

84. Hoffmann M, Wanko M, Strodel P, König PH, Frauenheim T, Schulten K, Thiel W, Tajkhorshid E, Elstner M (August 2006). "Color tuning in rhodopsins: the mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II". *Journal of the American Chemical Society*. **128** (33): 10808–18. doi:10.1021/ja062082i (https://doi.org/10.1021%2Fja062082i) . PMID 16910676 (https://pubmed.ncbi.nlm.nih.gov/16910676) .

85. Mendive-Tapia D, Mangaud E, Firmino T, de la Lande A, Desouter-Lecomte M, Meyer HD, Gatti F (2018). "Multidimensional Quantum Mechanical Modeling of Electron Transfer and Electronic Coherence in Plant Cryptochromes: The Role of Initial Bath Conditions". *J. Phys. Chem. B*. **122** (1): 126–136. doi:10.1021/acs.jpcb.7b10412 (https://doi.org/10.1021%2Facs.jpcb.7b10412) . PMID 29216421 (https://pubmed.ncbi.nlm.nih.gov/29216421) .

86. Strümpfer J, Schulten K (2012). "Open Quantum Dynamics Calculations with the Hierarchy Equations of Motion on Parallel Computers" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3480185) . *J. Chem. Theory Comput*. **8** (8): 2808–2816. doi:10.1021/ct3003833 (https://doi.org/10.1021%2Fct3003833) . PMC 3480185 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3480185) . PMID 23105920 (https://pubmed.ncbi.nlm.nih.gov/23105920) .

87. Scheraga HA, Khalili M, Liwo A (2007). "Protein-folding dynamics: overview of molecular simulation techniques". *Annual Review of Physical Chemistry*. **58**: 57–83. Bibcode:2007ARPC...58...57S (https://ui.adsabs.harvard.edu/abs/2007ARPC...58...57S) . doi:10.1146/annurev.physchem.58.032806.104614 (https://doi.org/10.1146%2Fannurev.physchem.58.032806.104614) . PMID 17034338 (https://pubmed.ncbi.nlm.nih.gov/17034338) .

88. Muñoz-Huerta, Rafael F.; Guevara-Gonzalez, Ramon G.; Contreras-Medina, Luis M.; Torres-Pacheco, Irineo; Prado-Olivarez, Juan; Ocampo-Velazquez, Rosalia V. (Aug 16, 2013). "A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3812630) . *Sensors*. **13** (8): 10823–10843. Bibcode:2013Senso..1310823M (https://ui.adsabs.harvard.edu/abs/2013Senso..1310823M) . doi:10.3390/s130810823 (https://doi.org/10.3390%2Fs130810823) . PMC 3812630 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3812630) . PMID 23959242 (https://pubmed.ncbi.nlm.nih.gov/23959242) .

89. Martin, P D; Malley, D F; Manning, G.; Fuller, L. (November 2002). "Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy". *Canadian Journal of Soil Science*. **82** (4): 413–422. doi:10.4141/S01-054 (https://doi.org/10.4141%2FS01-054) .

90. Brosnan JT (June 2003). "Interorgan amino acid transport and its regulation" (https://doi.org/10.1093%2Fjn%2F133.6.2068S) . *The Journal of Nutrition*. **133** (6 Suppl 1): 2068S–72S. doi:10.1093/jn/133.6.2068S (https://doi.org/10.1093%2Fjn%2F133.6.2068S) . PMID 12771367 (https://pubmed.ncbi.nlm.nih.gov/12771367) .

91. Watson TD (1998). "Diet and skin disease in dogs and cats" (https://doi.org/10.1093%2Fjn%2F128.12.2783S) . *The Journal of Nutrition*. **128** (12 Suppl): 2783S–89S. doi:10.1093/jn/128.12.2783S (https://doi.org/10.1093%2Fjn%2F128.12.2783S) . PMID 9868266 (https://pubmed.ncbi.nlm.nih.gov/9868266) .

92. Case LP, Daristotle L, Hayek MG, Raasch MF (2010). *Canine and Feline Nutrition-E-Book: A Resource for Companion Animal Professionals*. Elsevier Health Sciences.

Further reading

Textbooks

- Branden C, Tooze J (1999). *Introduction to Protein Structure*. New York: Garland Pub. ISBN 978-0-8153-2305-1.
- Murray RF, Harper HW, Granner DK, Mayes PA, Rodwell VW (2006). *Harper's Illustrated Biochemistry*. New York: Lange Medical Books/McGraw-Hill. ISBN 978-0-07-146197-9.
- Van Holde KE, Mathews CK (1996). *Biochemistry* (https://archive.org/details/biochemistry00math) . Menlo Park, California: Benjamin/Cummings Pub. Co., Inc. ISBN 978-0-8053-3931-4.

External links

Databases and projects

Protein
at Wikipedia's sister projects

Definitions from Wiktionary

Media from Commons

- NCBI Entrez Protein database (https://www.ncbi.nlm.nih.gov/sites/entrez?db=protein)
- NCBI Protein Structure database (https://www.ncbi.nlm.nih.gov/sites/entrez?db=structure)
- Human Protein Reference Database (https://web.archive.org/web/20060424071622/http://www.hprd.org/)
- Human Proteinpedia (https://web.archive.org/web/20070314135408/http://www.humanproteinpedia.org/)
- Folding@Home (Stanford University) (http://folding.stanford.edu/) Archived (https://web.archive.org/web/20120908075542/http://folding.stanford.edu/English/HomePage) 2012-09-08 at the Wayback Machine
- Protein Databank in Europe (http://www.pdbe.org/) (see also PDBeQuips (https://archive.today/20130727184433/http://www.pdbe.org/quips) , short articles and tutorials on interesting

PDB structures)

- Research Collaboratory for Structural Bioinformatics (http://www.rcsb.org/) (see also Molecule of the Month (http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/index.html) Archived (https://web.archive.org/web/20200724151351/https://www.rcsb.org/pdb/static.do?p=education_discussion%2Fmolecule_of_the_month%2Findex.html) 2020-07-24 at the Wayback Machine, presenting short accounts on selected proteins from the PDB)
- Proteopedia – Life in 3D (http://www.proteopedia.org/) : rotatable, zoomable 3D model with wiki annotations for every known protein molecular structure.
- UniProt the Universal Protein Resource (https://web.archive.org/web/20080608183902/http://www.expasy.uniprot.org/)

Tutorials and educational websites

- "An Introduction to Proteins" (https://web.stanford.edu/group/hopes/cgi-bin/hopes_test/an-introduction-to-proteins/) from HOPES (Huntington's Disease Outreach Project for Education at Stanford)
- Proteins: Biogenesis to Degradation – The Virtual Library of Biochemistry and Cell Biology (https://web.archive.org/web/20050219090405/http://www.biochemweb.org/proteins.shtml)

Portals: **Biology** **Technology** **Medicine** **Chemistry** **Food** **Ecology** **Environment** **Science** **Evolutionary biology**